Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 25C (2017) 2400–2411

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016

Spatial distribution of urban trips in recently expanded Surat city through Fuzzy Logic with various clustering Techniques: A case study of typical metropolitan city in India

Salini P S[1], A. Kedia[2], S.Dhulipala[3], Krishna Saw[4], B.K. Katti[5]

[1,2,3] *P.G. Student, Civil Engineering Department, SVNIT, Surat – 395007, Gujarat, India*
[4]*Institute of Urban Transport, New Delhi, India*
[5]*Visiting Professor, Civil Engineering Department, SVNIT, Surat – 395007, Gujarat, India*

Abstract

Trip distribution finds prime place after trip generation in sequential modelling of travel demand to cover the spatial dimensions in a geographical area, to reflect on trip length and frequency. It provides the basis for strategic land use and transport infrastructure development both at local and regional levels. Trip distribution problems in the real world are quite complex with association of uncertainty in the decision making and therefore calls for an unorthodox approach to deal with the concerned issue. Soft computing technique - Fuzzy Logic (FL) is believed to be capable of addressing the uncertainty lying in the travellers' behaviour and has been sought to develop realistic behavioural models in the recent years. FL takes into account linguistic variables and is based on simple and logical "IF-THEN" rules which closely resemble human thought process. Fuzzy Logic based trip distribution models are developed employing Fuzzy C-mean (FCM) clustering, and are compared for their performance with the *Genfis* based approach, where a Sugeno-type Fuzzy Inference System (FIS) is generated using Subtractive clustering. Surat, a fast growing metropolitan city in India is considered to realize the study. The models developed here, find applications in strategic land-use and transport planning for developing Indian cities.

*Keywords:*Travel demand estimation, Trip distribution, Fuzzy Logic, Fuzzy C mean clustering, Subtractive clustering, Genfis

2352-1465 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.
10.1016/j.trpro.2017.05.245

1. Introduction

Four stage modelling process is the commonly employed method for forecasting travel demand, involving phases such as trip generation, trip distribution, mode choice and traffic assignment. Trip distribution analysis plays vital role as it deals with the destination choice phase. Conventional trip distribution models which has major limitation that the input variables need to have crisp values and hence should be measured accurately, which consumes lot of time and resources. Moreover, decision of trip maker for choosing a destination involves human approximations which are not precisely captured by these conventional models. This can be overcome by using artificial intelligence techniques like fuzzy logic for modeling trip distribution. Fuzzy logic try to harness the human knowledge which is often guided by approximations by accepting input values in linguistic terms. The fuzzy rule base comprises of several IF-THEN rules which closely resemble human knowledge and decision-making. Thus to overcome limitation of traditional model, application of fuzzy logic seems to be most appropriate to consider ambiguity, imprecision and vagueness in influencing variables. Fuzzy logic model in contrast to the conventional model takes into account highly non-linear relationship between input and output variables.

Clustering is the effective tool for natural grouping of data from a large data set, and allows concise representation of characteristics of the data. Given a data set, the aim of clustering is to partition it into different groups (clusters) so that the members in the same group are of similar nature, whereas members of different groups are dissimilar. While clustering, various similarity measures can be considered, one of the most commonly used is distance between data samples. The Fuzzy Inference System (FIS) is a simple way to construct systems models without using complex analytical equations. The Fuzzy C-mean algorithm and Subtractive clustering algorithm are implemented to find the number and the position of clusters for a set of highly non-linear data. In the present study, different FIS are created using Fuzzy C mean clustering as well as Subtractive clustering. Fuzzy C-mean clustering provides a method that shows how to group data points that populate some multidimensional space into a specific number of different clusters. As for subtractive clustering, the radii parameter is changed to obtain different number of clusters. Generally, increasing the number of generated cluster yields an improvement in result. For subtractive clustering the radii parameters are tuned. This automatically generates the number of clusters. The fuzzy models developed with the help of these two clustering techniques are compared here to find out the efficiency of the clustering systems and thereby arriving at realistic trip distribution models.

2. Literature Review

Fuzzy Logic (FL) concept was introduced by Zadeh [1]. A method of generating fuzzy rules from the numerical data was demonstrated, which facilitated the combination of numerical as well as linguistic information into a common assembly – fuzzy rule base [2]. A pioneering fuzzy logic approach to trip distribution modeling was put forward by estimation of air passenger flows among selected major industrial cities and tourist resorts using known productions and attractions as inputs [3]. Later, another study by the same research group showed improved performance with a hybrid Genetic Fuzzy Rule Based System (GFRBS) design [4]. Trip generation and distribution modelling using fuzzy logic was further attempted, considering trip generation at country level, whereas trip distribution between origin country and destination countries (country-pair level) and provided empirical evidence relating to successful use of fuzzy logic as a non-traditional technique [5]. Discretionary trips being flexible than compulsory trips, are characterized by uncertainty in decision making, modelling the behavior towards such trips is observed to be addressed well by fuzzy logic based approach [6]. Several studies advocated the use of Fuzzy Logic as a tool to address the uncertainty in the people's expressions and imprecision in their perception towards travel attributes [7]. Also, Fuzzy logic systems could be treated as universal approximators, which mean that a fuzzy logic system can uniformly approximate any real continuous non-linear function to an arbitrary degree of accuracy [8, 9]. Fuzzy Inference System (FIS) can be of two types, viz. Mamdani and Sugeno FIS. Mamdani fuzzy inference system finds application in wide variety of disciplines due to intuitive and interpretable nature of its fuzzy rule base [10, 11].In Mamdani-type the consequent of each rule is a fuzzy set, whereas in Sugeno-type the consequent is a function of input variables. Due to this difference, the inference mechanism of determining the output of the system in both the categories varies to some extent. The basis of selection of optimum Membership Functions (MFs) for fuzzifying the input variables has always been a matter of discussion. For the same, Clustering techniques have been employed to arrive at the shape of membership functions in few studies. Two most famous clustering techniques: fuzzy c-means algorithm and subtractive clustering algorithm were compared based on validity measurement of their clustering results for highly non-linear functions and the latter technique was observed to outperform the other one [12]. Another recent application of fuzzy logic – subtractive clustering technique was demonstrated at different stages of four-stage travel demand modelling [13].The present study evaluates and compare the efficiencies of the Fuzzy trip distribution models generated through Fuzzy C-mean clustering and Subtractive clustering techniques taking a case of Indian city, Surat. Apart from Fuzzy Logic, other Soft Computing techniques such as Artificial Neural Network (ANN) and Genetic Algorithm (GA) are also sought by researchers for solving problems pertaining to transportation planning in the recent years, and scope lies for further improvement in performance of models.

3. Data collection

Surat is Gujarat's second-largest city with a population of 4.6 million (Census 2011). It is the eighth largest city and the ninth largest metropolitan area of India. Surat consists of seven administrative zones, and West zone on the right bank of river Tapti is considered in the study, analysis and development of the micro level distribution model which is having a population of 4.3 lakhs and spread over an area of 51.28 sq.km. Variyav portion on the north side of the zone is left over owing to its meagre population having mainly agricultural land-use. As such, it is predominantly a residential zone. The study is concentrated over a part of West zone having an area of 28 sq. km encircling 4 lakhs population. The study zone is further divided into six sub-zones, namely Adajan, Rander, Jahangirpura, Jahangirbad, Pal and Palanpur to carry out model building and analysis. Land-use of the area has great influence on attracting trips. The study area is having a mixed land-use, but its share of industry and commercial activity is very meagre. Residential as well as an agricultural area finds the major land-use share in Jahangirpura and Jahangirbad. Educational institutes of pre and primary schools are available in every sub-zone.

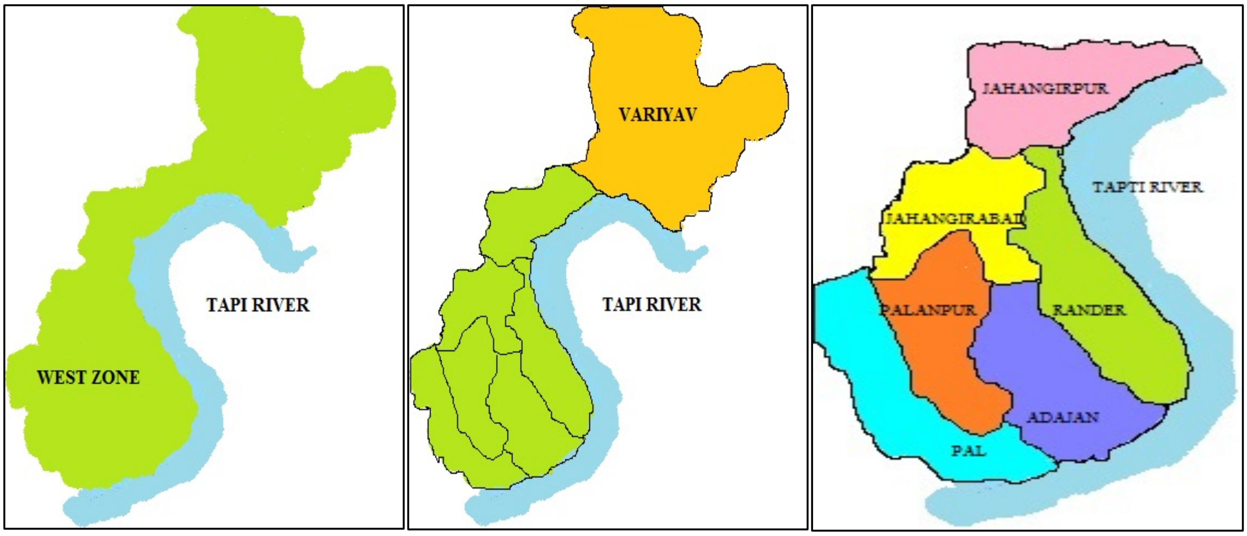

Figure 1. Study area

The necessary data base for the present study is developed by inventory surveys (secondary data) as well as primary survey in the form of home interviews. The inventory data helps in designing the field interviews, in terms of the sample size, survey technique and sampling methodology. The inventory data at present is built up on the collection of study relevant data from various sources as SMC (Surat Municipal Corporation) and SUDA (Surat Urban Development Authority). The demographic profile of the sub-zones is presented in Table 1 to provide particulars on area, density, population and households.

Table.1. Demographic profile of study zone (2011)

Sub-Zone	Area(km^2)	Population	No. of HHs	Density
Adajan	6.73	1,96,850	45,993	29,249.6
Rander	5.12	1,14,586	23,290	22,380.1
Jahangirpura (JP)	3.62	27,807	5,584	7,689.9
Jahangirbad (JB)	4.16	6,221	1,567	1,495.4
Pal	6.05	36,108	9,118	5,973.2
Palanpur (PP)	3.01	23,492	5,932	7,809.8
Total	28.68	4,05,064	91,484	74,598.2

The information pertaining to existing conditions with respect to urban land-use, transport system, economic activity profile, travel pattern, and socio economic characteristics of the study area is vital for any meaningful urban transportation planning exercise. The efficiency of model developed depends heavily on the calibration data used. Home interview surveys are carried out from nearly 400 households to build the data base for the study using designed questionnaire. The data focus was on trip origin, trip destinations and trip purpose among the six sub-zones for intra zonal analysis. Trips carried from each sub-zone outside the West zone across the river to the other city area were also noted. 70% of the data is used for the development of the model and rest 30% is used for validation of the models.

The survey data shows that nearly 34% are working members in the study zone, whereas student population is almost 28%. The percentage of non-working members is also noticeable at 33%. Therefore, the influence of non-working members may affect the trip making behaviour. From the survey data analysis, it is observed that 20% of the household samples belong to Low Income Group (LIG), 41% belongs to Middle Income Group (MIG) and 39% have its place in High Income Group (HIG).

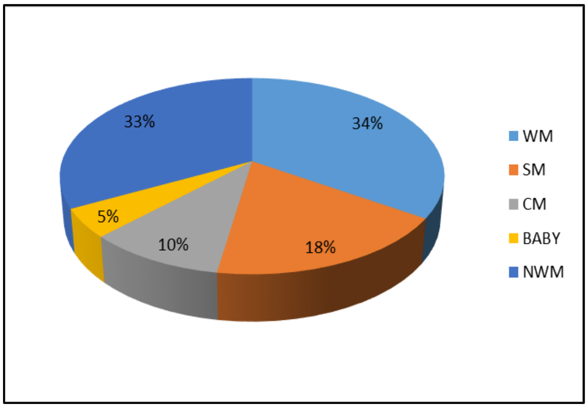

Figure 2 (a).Household structure of West zone

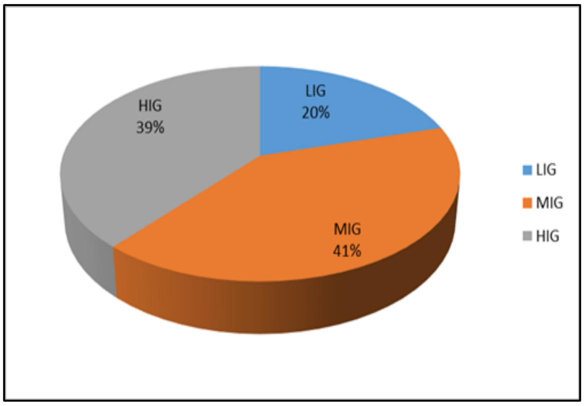

Figure 2(b).Income distribution in West zone

4. Methodology

4.1 Model inputs

Major inputs generally qualifying for the models are Trip generation at i^{th} zone, Trip attraction by j^{th} zone and zonal separation measured in terms of distance/ travel time/travel cost. These parameters influence the trips between two zones. In the present work, Trip generation, Trip attraction and the Trip distance (km) between two subzones i-j of the study zone forms the set of predictor variables. Sub zonal trip generations for the entire six sub-zones depend on Trips per capita per day (TPCD) and sub-zonal population. TPCD values are computed from the survey database. Total trips generated and attracted is 800 numbers inclusive of all income status. Trips going outside the study area are not considered for the development of the model. All the mandatory as well as discretionary trips are considered here. Total trip production of LIG, MIG and HIG are given in Table 2.

Table 2.Zonal Productions/Attractions*

Sub-Zone	P_i	A_i
Adajan	264	364
Rander	117	121
Pal	34	58
Palanpur	36	103
Jahangirpura	192	74
Jahangirbad	157	80
Total	800	800

*From the survey data

Inter sub-zonal distances in km for all the six sub-zones are mentioned in the following Table.3. Average distance based on survey data is considered for the intra sub-zonal trips. The maximum distance of 6.71 km is between Jahangirpura and Pal. The minimum distance of 1.30 km is observed between Palanpur and Jahangirbad. Lower distances are observed in a diagonal way for the intra sub-zonal trips. All the distances are measured between the zonal centroid.

Table.3. Inter sub-zonal distance matrix (km)

Sub zones	Adajan	Rander	Pal	Palanpur	Jahangirpura	Jahangirbad
Adajan	1.50*	2.65	2.42	3.60	6.10	3.22
Rander	2.65	1.20	3.33	2.60	3.60	1.41
Pal	2.42	3.33	1.00	2.21	6.71	2.80
Palanpur	3.60	2.60	2.21	0.70	5.07	1.30
Jahangirpura	6.10	3.60	6.71	5.07	1.50	4.01
Jahangirabad	3.22	1.41	2.80	1.30	4.01	1.00

**intra sub-zone distance= i*i is considered as average trip length in km on observed data.*

The necessary information base for the present study is developed by home interview surveys. Nearly 400 household data were collected by designed questionnaire survey. Trip Origin, Trip destination and trip purposes are collected for the development of the model. Based on the household income the whole data were divided into three income categories LIG,MIG and HIG having income ranges less than or equal to Rs. 15000 ,between 15000 and 30000 and more than 30000 respectively.

Distance Impedance Factor (DIF) **'α'** in Gravity model has been calibrated on existing trip

$$\frac{Tij = Pi\,Pj}{Dij^{\propto}} K$$

Distribution pattern for three categories of income groups to find **'α'** equal to 1.60, 1.64 and 1.74 for HIG, MIG and LIG. These **'α'** are further employed in Fuzzy Model development to define accessibility levels. Accessibility is $1/D_{ij}^{\alpha}$.

4.2Fuzzy Logic model using Fuzzy C-mean clustering

The fuzzy model is developed using the fuzzy tool box available in MATLAB. The fuzzy model structure comprises of three basic stages as shown in Figure 3.

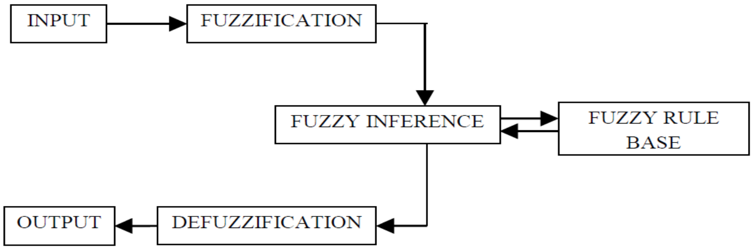

Figure 3. Fuzzy Logic Structure

First stage is Fuzzification, the process where the crisp quantities are converted into linguistic fuzzy set. A membership function is a mathematical expression that deals with fuzziness of a fuzzy set. Fuzzy inference system (FIS) is the second and important phase of fuzzy logic system. The FIS formulates suitable “IF-THEN” rules based upon which the decision process takes place. Third is the Defuzzification phase where the aggregated fuzzy output is converted into a crisp value ready for application.

Here, the shapes of the membership functions for the input variables, viz. trip productions, trip attractions and distances between the sub-zones, and output variable trip interchanges have been obtained by carrying out the fuzzy C-mean clustering in MATLAB. All the input variables are divided into three fuzzy sets, whereas the output variable into five fuzzy sets. Several parameters must be specified in order to carry FCM algorithm such as the *number of clusters* c, *fuzziness exponent* m, *termination tolerance* ε, and the *norm-inducing matrix* A. Finally, the fuzzy partition matrix U must be initialized. The number of clusters c is the most important parameter. When clustering real data without any prior information about the structures in the data, one usually has to make assumptions about the number of underlying clusters. The chosen clustering algorithm then searches for c clusters, regardless of whether they are really present in the data or not. The validity measure approach and iterative merging or insertion of clusters approach are the main approaches used to determine the appropriate number of clusters in data. The shapes and ranges of the membership functions of the considered input variables *trip productions*, *trip attractions* and *distances* between sub-zones obtained through fuzzy C-means clustering, for the three income groups are shown in Figure 4, 5 and 6 respectively, and the same for output is shown in Figure 7 which is obtained from the fuzzy model. In Figure 4, 5, 6 and 7, X axis represents Trip Productions, Trip attractions, Accessibility and Trip interchanges respectively. Y axis represents degree of membership.

- Trip Productions

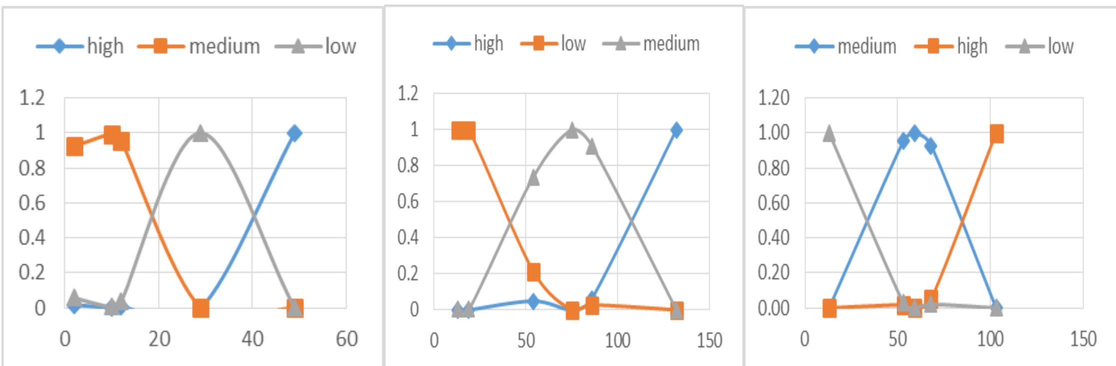

Figure 4 (a) LIG **Figure 4 (b) MIG** **Figure 4 (c) HIG**

- Trip Attractions

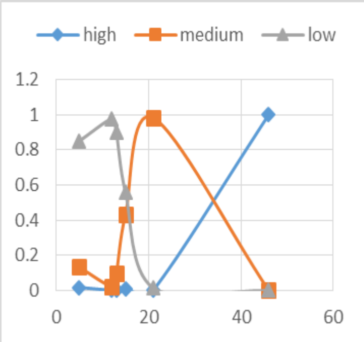

Figure 5 (a) LIG

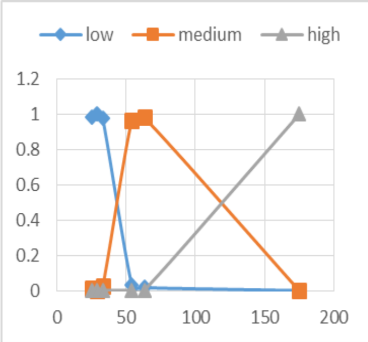

Figure 5 (b) MIG

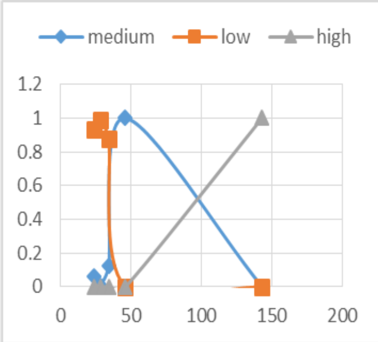

Figure 5 (c) HIG

- Accessibility

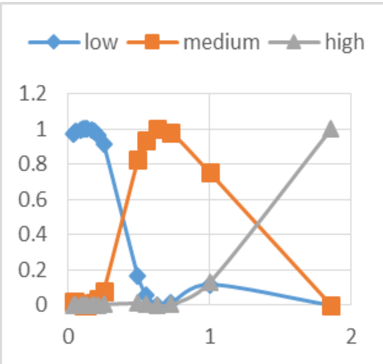

Figure 6 (a) LIG

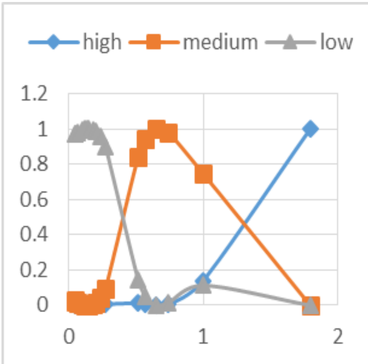

Figure 6 (b) MIG

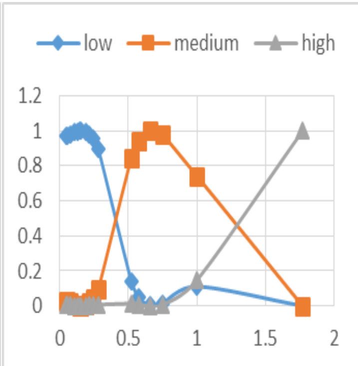

Figure 6 (c) HIG

- Trip Interchanges

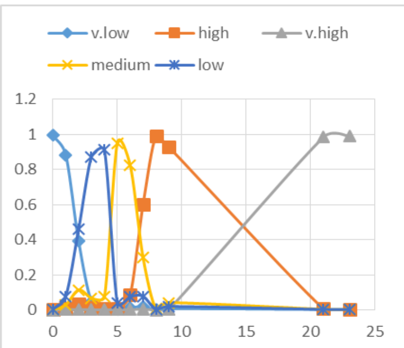

Figure 7 (a) LIG

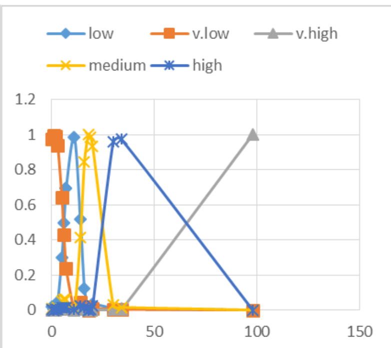

Figure 7 (b) MIG

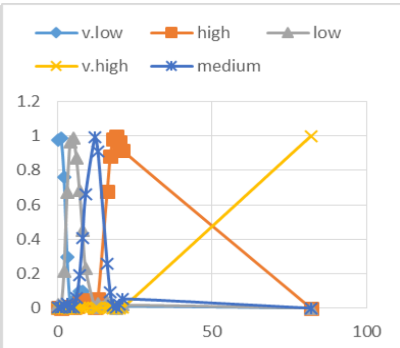

Figure 7 (c) HIG

Fuzzy inference mechanism works on the principle of simple and logical If-then rules. Total 27 numbers of “IF-THEN” rules are formulated here based on the logical thinking and general experience. A typical set of rules can be as follows:

IF<zonal productions are high> AND <attractions are high> AND <accessibility is less>, THEN <the number of trip interchanges are very high>.

IF<zonal productions are low> AND <attractions are medium> AND <accessibility is less>, THEN <the number of trip interchanges are low>.

IF<zonal productions are high> AND <attractions are medium> AND <accessibility is medium>, THEN <the number of trip interchanges are medium>.

IF<zonal productions are medium> AND <attractions are low> AND <accessibility is high>, THEN <the number of trip interchanges are medium>.

The fuzzy output of the fuzzy inference system cannot be used directly for application purpose and hence the defuzzification of the aggregated output fuzzy sets is carried by using Centroid method, to convert the linguistic output into crisp format.

4.3 Fuzzy Logic model using subtractive clustering

Fuzzy logic based modelling is attempted here, using subtractive clustering. The *genfis2* function in MATLAB, which uses subtractive clustering algorithm for identifying cluster centres and radii, was chosen to build the fuzzy inference system. The default input membership functions for the first-order Sugeno-type fuzzy inference system configured using genfis2 functions are Gaussian 'bell-shaped' type. Output variable considered is linear. The g*enfis2* generates a Sugeno-type FIS structure using subtractive clustering and requires separate sets of input and output data as input arguments. It accomplishes this by extracting a set of rules that models the data behaviour. The rule extraction method first uses the *subclust* function to determine the number of rules and antecedent membership functions and then uses linear least squares estimation to determine each rule's consequent equations. This function returns a FIS structure that contains a set of fuzzy rules to cover the feature space. The shapes and ranges of the membership functions of the considered input variables obtained through subtractive clustering for the three income groups LIG, MIG and HIG are shown in Figure 8, 9 and 10 respectively. All the membership functions arrived at through subtractive clustering are observed to have 'bell shape'.

- Input Membership Functions for LIG

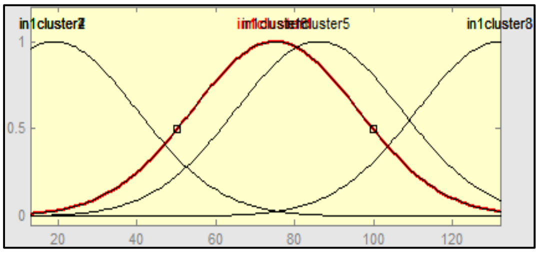

Figure 8(a) Trip production (LIG)

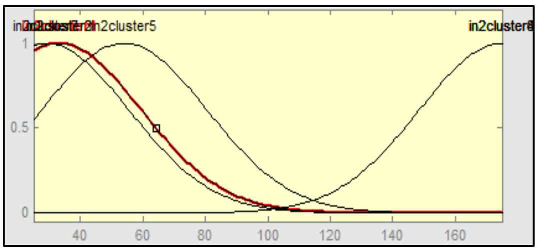

Figure 8(b) Trip attraction (LIG)

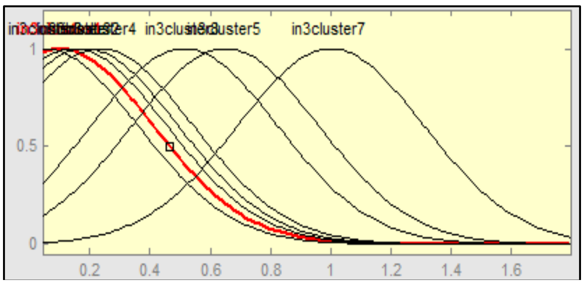

Figure 8(c) Accessibility (LIG)

- Input Membership Functions for MIG

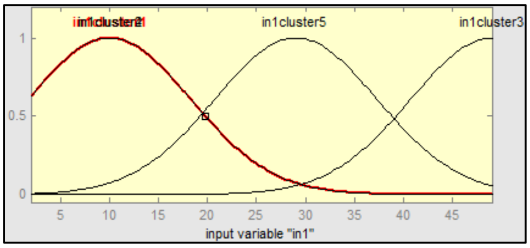

Figure 9(a) Trip production (MIG)

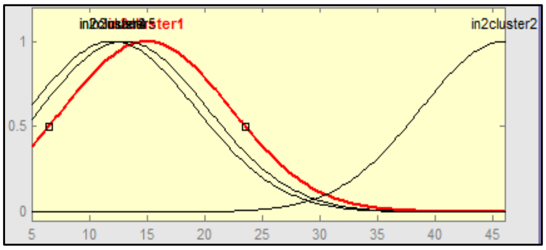

Figure 9(b) Trip attraction (MIG)

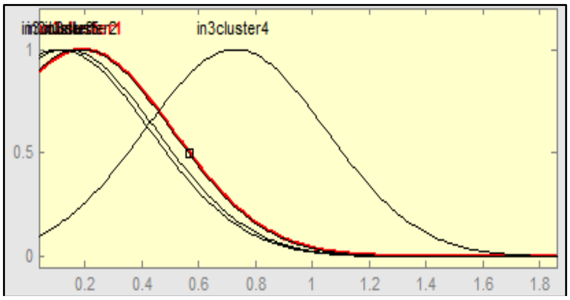

Figure 9(c) Accessibility (MIG)

- Input Membership Functions for HIG

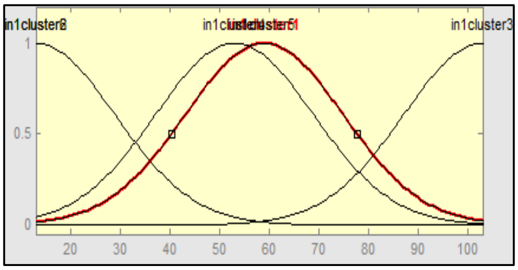

Figure 10(a) Trip production (HIG)

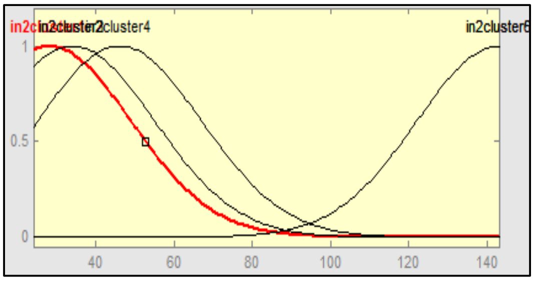

Figure 10(b) Trip attraction (HIG)

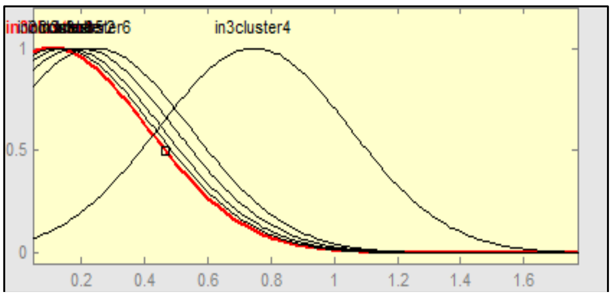

Figure 10(c) Accessibility (HIG)

Membership functions for the HIG trip distribution models are obtained as above. Output of the model i.e. Trip interchanges between the zones is obtained by evaluating the FIS file which is obtained from Genfis2.

5. Results and Discussions

Fuzzy Logic trip distribution models are developed for the three socio-economic classes of Indian society, one using Fuzzy C-mean clustering and other using the subtractive clustering method. Prediction accuracy of the models by the two approaches is compared with the help of Root mean squared error. Also, these two approaches are observed across the conventional Gravity model. Three different Gravity models were developed by adopting accessibility influencing index value (α) as 1.74, 1.64 and 1.60 for LIG, MIG and HIG respectively. RMSE values for the models are given in Table 4.

Table 4.RMSE Values

Income groups	Gravity model	Fuzzy Model (Fuzzy C-means clustering)	Fuzzy Model (Subtractive clustering)
LIG	2.65	2.42	1.248
MIG	5.12	4.74	2.205
HIG	3.95	2.78	2.232

From the above RMSE observations, it is clear that both type of fuzzy models are performing better than the conventional Gravity model, this is due to the uncertainty addressing ability of fuzzy model. Moreover, it can be inferred that Fuzzy modelling approach with subtractive clustering outweighs the fuzzy model with Fuzzy C mean clustering and the conventional gravity based technique.

These models are developed based on the samples obtained from the interview survey data, an attempt has been made here to forecast the trips for the entire population of the west zone. Using the expansion ratio, the trip interchanges obtained from the survey data are projected for whole population. A_i/P_i ratio is found out for analyzing the performance of models and the values are given below in the Table5.

Table 5. Observed and Predicted A_i/P_i ratios

Sub zones	**Observed A_i/P_i**	**Modelled A_i/P_i**		
		Gravity Model	Fuzzy Model (Fuzzy C-mean clustering)	Fuzzy Model (Subtractive clustering)
Adajan	1.19	1.09	1.06	1.21
Rander	0.65	0.67	0.71	0.58
Pal	1.15	0.82	1.20	1.15
Palanpur	5.99	5.54	5.71	5.33
Jahangirpura	0.44	0.57	0.61	0.37
Jahangirbad	0.88	1.67	1.21	1.35
Chi-Squared Value		0.578	0.171	0.164

Lower Chi-square value has been observed for the fuzzy model with Subtractive clustering indicating it as a better operating model than other two models. Surface viewer window for the fuzzy model with subtractive clustering is given below in Figure 11. It shows Trip production and Trip attraction are directly proportional to trip interchanges.

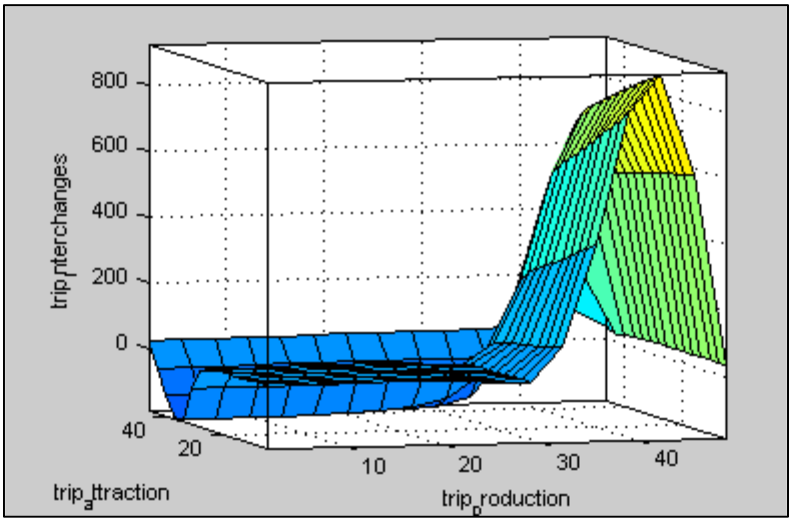

Figure 11 Surface plots for predicted model

6. Conclusions

Among the four stages of sequential travel demand modelling process, *trip distribution* bears huge importance as it deals with the destination choice phase which provides the town and transport planners with the intensities of future trip interchanges among various O-D pairs. This guides in strategic planning for remedial measures to avoid the demand-supply gap in the future years. Soft computing models can address the uncertainty associated with the decision making process of the travelers'. From the RMSE and A_i/P_i observations, it can be inferred that soft computing technique - Fuzzy Logic based approach is giving satisfactory result than the conventional Gravity model, and supports the applicability of Fuzzy Logic in Trip distribution problems effectively. Out of the two Fuzzy models developed, model with Subtractive clustering is observed to perform better compared to the one with Fuzzy c mean clustering approach. A training algorithm is needed to accurately generate models using FCM. However, subtractive clustering does not need training algorithm. FCM has inconsistence problem such that, different running of the FCM yield different results as the algorithm will choose an arbitrary υ matrix each time. On the other hand, subtractive clustering algorithm produces consistent results.

Acknowledgements

We thank all the respondents of the Household Interview Surveys conducted in Surat city for giving their precious feedbacks, without which it would not have been possible to conduct this study.

References

1. Zadeh, L. A. 1965. Fuzzy sets, Information and Control 8(3): 338–353. http://dx.doi.org/10.1016/S0019-9958(65)90241-X
2. Wang, L.X. and Mendel, J.M. (1992) 'Generating fuzzy rules by learning from examples', *IEEE Transactions on Systems*, *Man and Cybernetics* 22, 1414–1427.
3. Kalić, M.; Teodorović, D. 1996. Solving the trip distribution problem by fuzzy rules generated by learning from examples, in Proceedings of the XXIII Yugoslav Symposium on Operations Research, 1–5 October 1996, Zlatibor, Yugoslavia, 777–780. (in Serbian).
4. Kalic. M, Teodorovic, D. (2003) Trip distribution modelling using fuzzy logic and a genetic algorithm, *Transportation Planning and Technology*,Vol.26(3),pp.213-238.

5. Kalic M, Kuljanin J & Dozic S, (2012) Air travel demand fuzzy modelling: trip generation and trip distribution.In: *Online Conference on Soft Computing in Industrial Applications Anywhere on Earth*, 10-12 Dec.
6. Y.Shafahi (2008) Fuzzy Trip Distribution Models for Discretionary Trips-*Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems.*
7. Taylor, P., & Yaldi, G. 2010. Examining the possibility of fuzzy set theory application in travel demand modelling. *Journal of Eastern Asia Society for Transportation Studies*.
8. Wang, L.X. (1992) 'Fuzzy systems are universal approximators', *Proc. IEEE International Conference on Fuzzy Systems,* San Diego, 1163–1170.
9. Wang, L.X. (1994) *Adaptive Fuzzy Systems and Control: Design and Stability Analysis* (PTR Prentice Hall, Englewood Cliffs, NJ).
10. Mendel, J. M. 2001. Uncertain Rule-Based Fuzzy Logic Systems:Introduction and New Directions. Prentice Hall. 576 p.
11. Jassbi, J.; Alavi, S. H.; Serra, P. J. A.; Ribeiro, R. A. 2007. Transformationof a Mamdani FIS to first order Sugeno FIS, inFUZZ-IEEE 2007: IEEE International Fuzzy Systems Conference2007, 23–26 July 2007, London, 1–6.http://dx.doi.org/10.1109/FUZZY.2007.4295331
12. Bataineh, K. M.; Naji, M.; Saqer, M. 2011. A comparison studybetween various fuzzy clustering algorithms, Jordan Journalof Mechanical and Industrial Engineering 5(4): 335–343.
13. Pulugurta, S.; Errampalli, M.; Kayitha, R. 2014. Fuzzy logicbasedtravel demand model to simulate public transportpolicies, Journal of Urban Planning and Development.http://dx.doi.org/10.1061/(ASCE)UP.1943-5444.0000261.